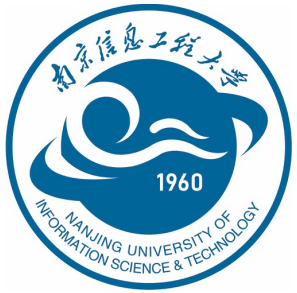
From Laboratory to Real World: A New Benchmark Towards Privacy-Preserved Visible-Infrared Person Re-Identification

Yan Jiang[1,2], Hao Yu[2], Xu Cheng[1*], Haoyu Chen[2], Zhaodong Sun[1,2], Guoying Zhao[2]

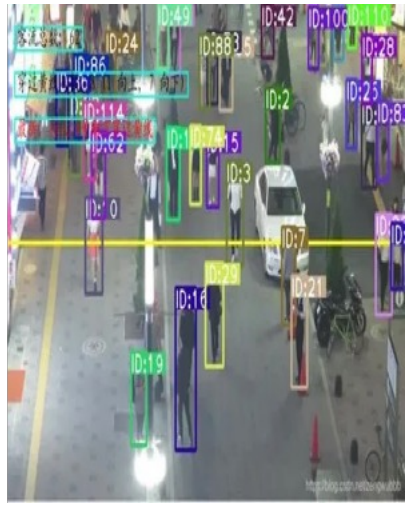
1 School of Computer Science, Nanjing University of Information Science and Technology
2 Center for Machine Vision and Signal Analysis, University of Oulu



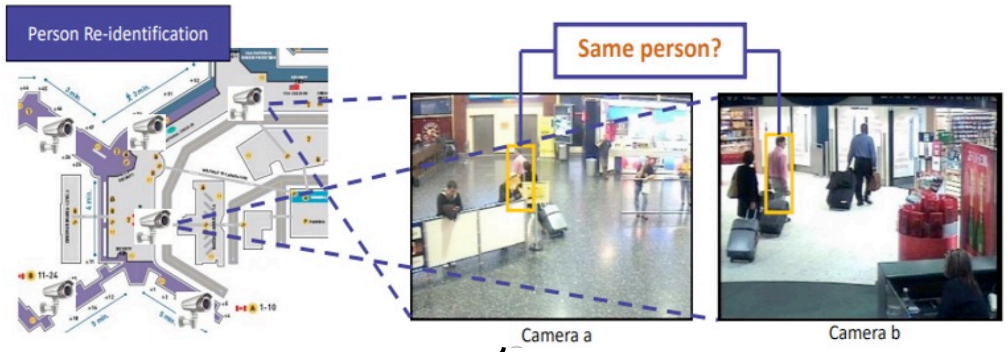
What is Person Re-Identification?

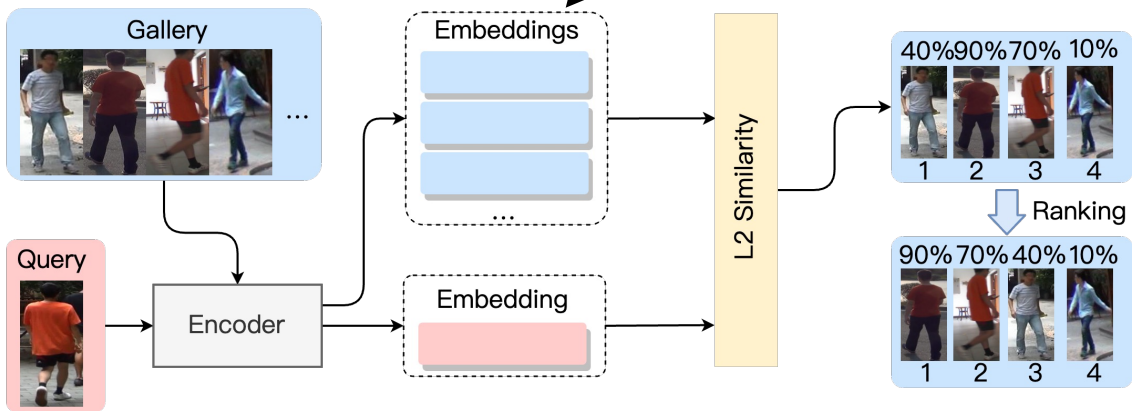
Surveillance systems

Criminal investigation

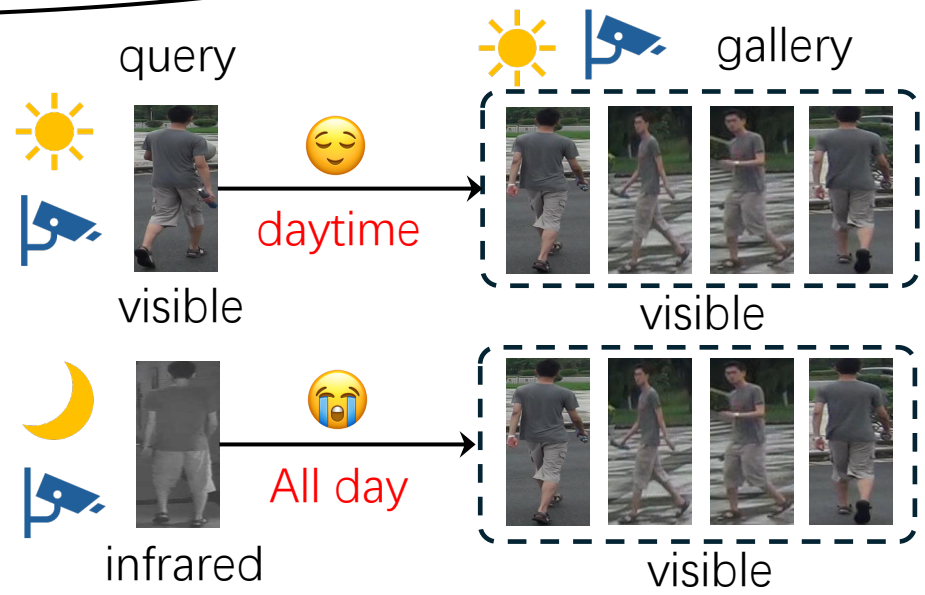


- Match individuals across **different cameras** based on their visual appearance.
- Play an important role in public safety

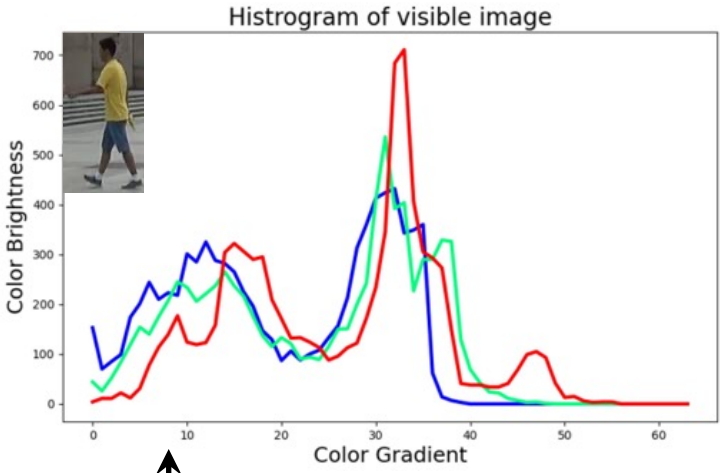




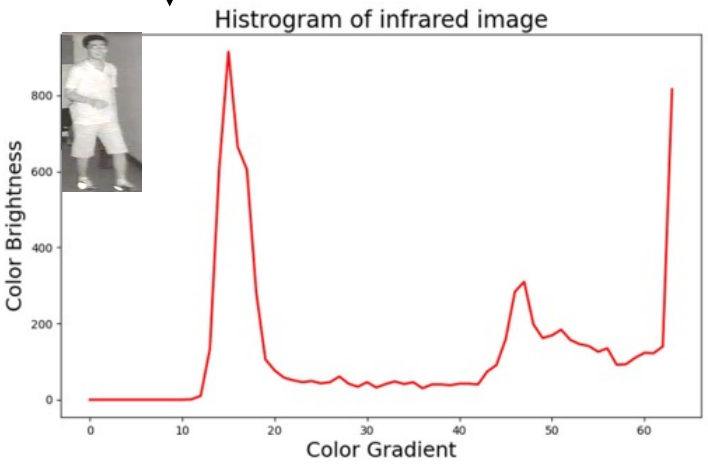
The higher the probability, the more likely it is that they are the same person.



Why is it hard to retrieval infrared images with visible images?



Huge gap



R, G, B
Wavelength:400nm-700nm
Abundant color

R
Wavelength:700nm-1mm
Without color

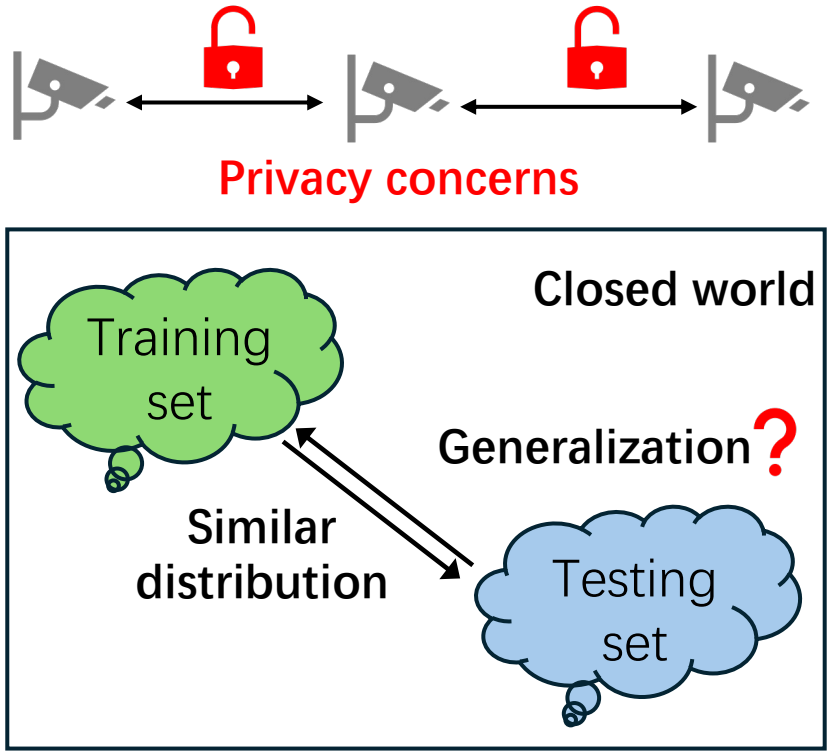
Further challenges

Despite cross-modality gap, alongside intra-modality gap like person Re-ID

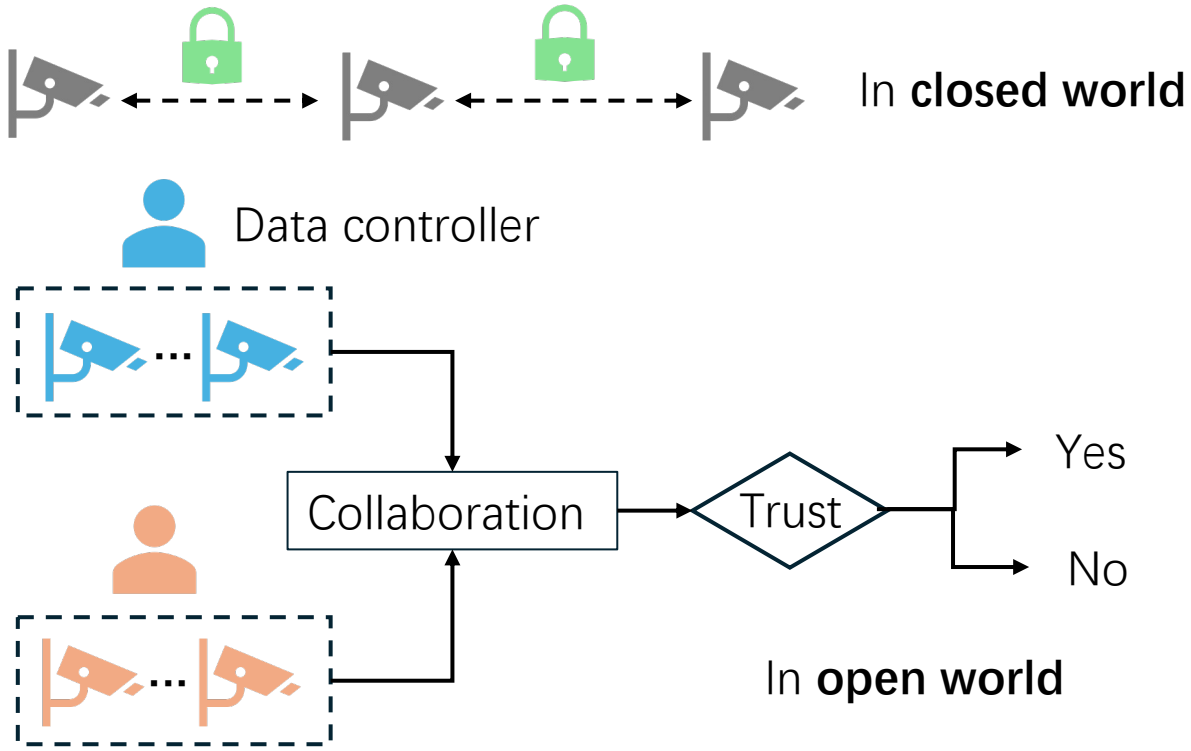
Motivation

Although existing VI-ReID methods achieve encouraging achievements, their development in real-world scenarios remains limited.

Existing settings:



Our L2RW settings (simulate real world)

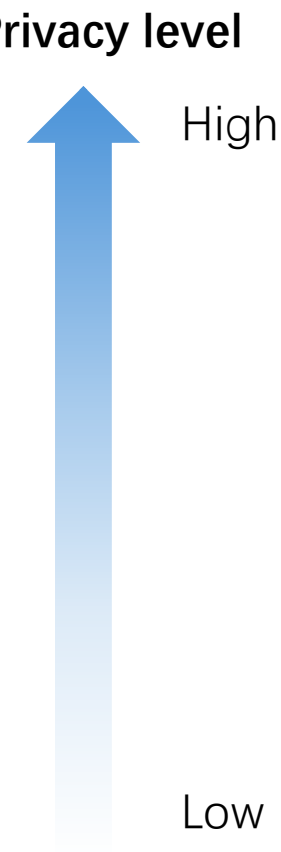


Designed Protocols in Our L2RW Benchmark

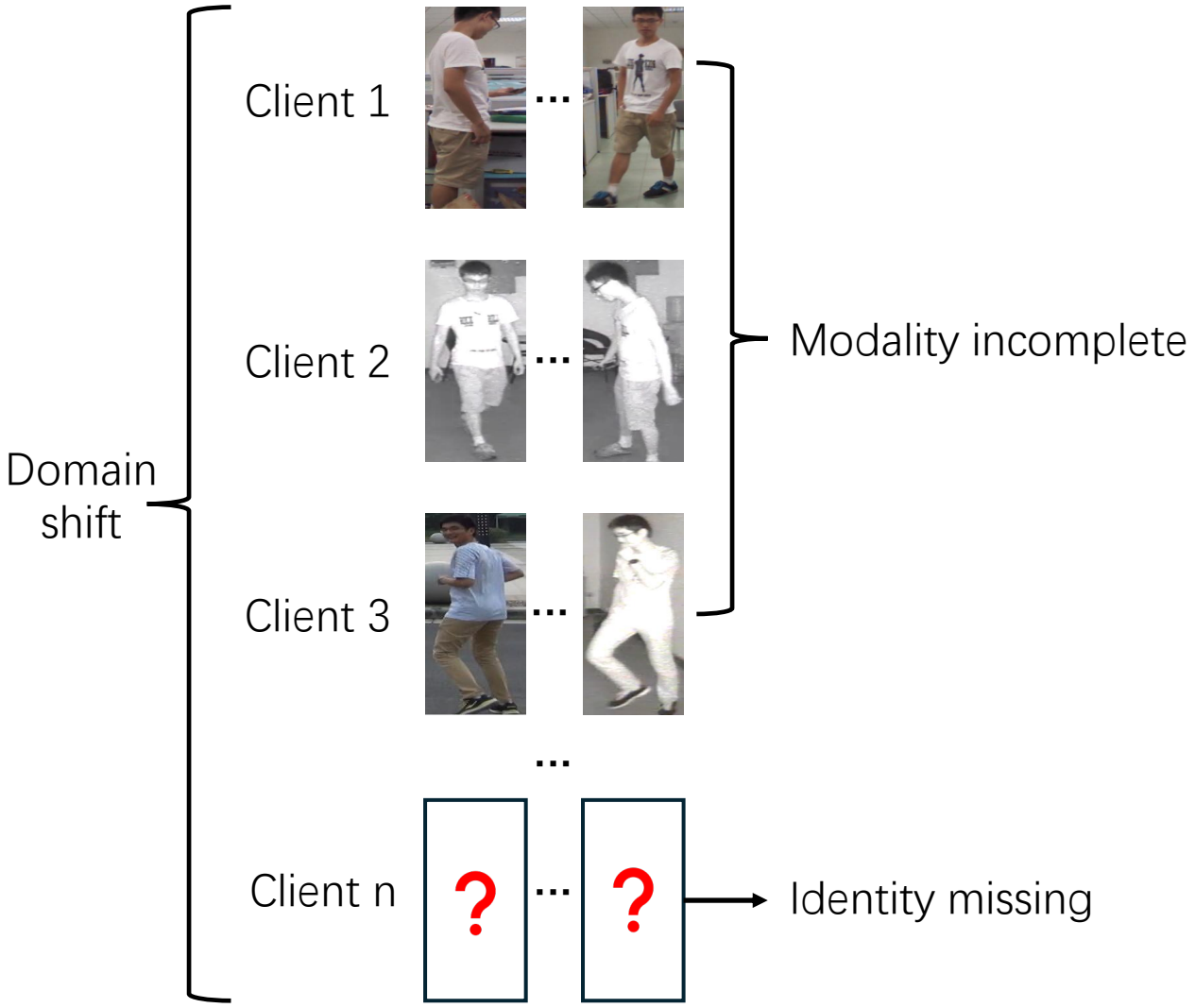
- **Camera Independence (CI) :** all camera data remains isolated, which simulates the most strict setting where no data sharing occurs across every single device.

- **Entity Independence (EI):** Data from different entities is kept isolated during training, and the model is trained using a decentralized way. Then evaluate the trained model on an unseen entity to assess its generalization ability.

- **Entity Sharing (ES):** Data from different entities is shared during training, and the model is evaluated on an unseen entity to assess its generalization ability.



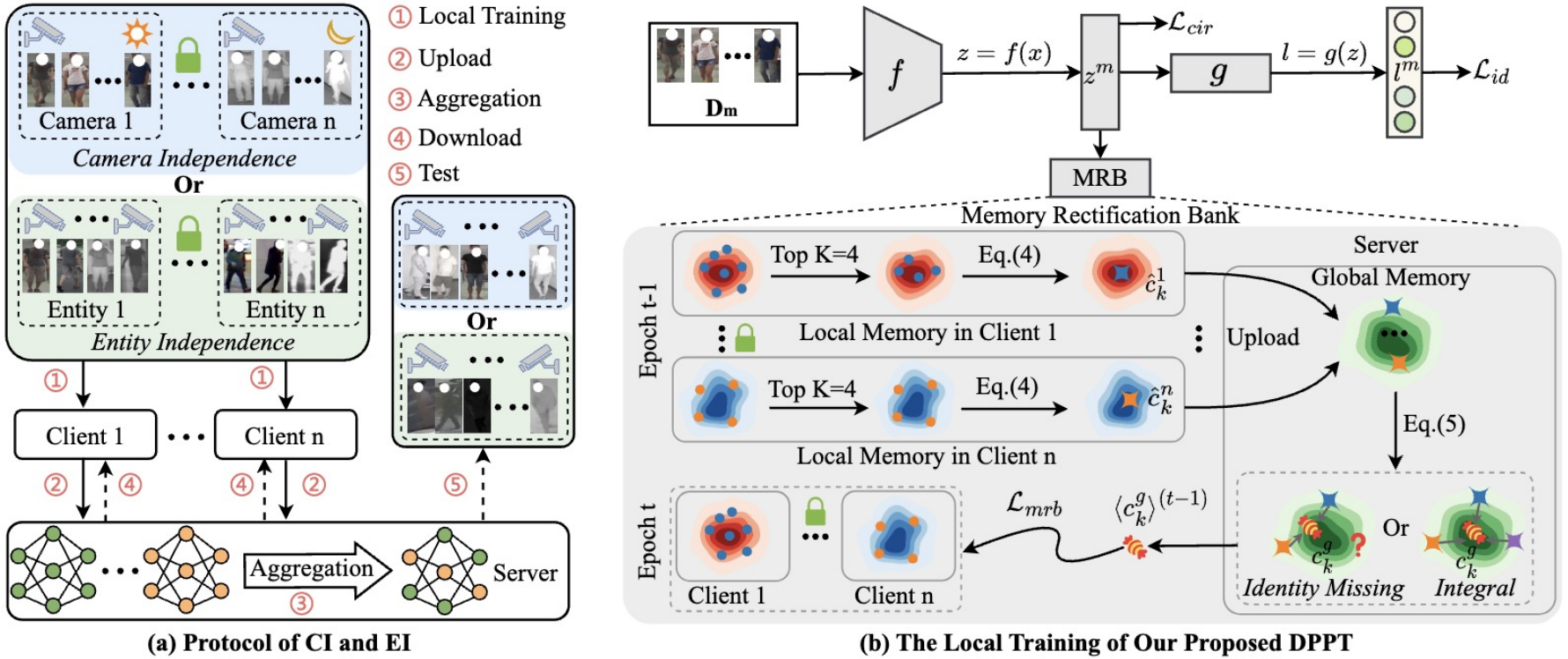
New Issues Encountered Under Data Independence (CI & EI)



- **Domain shift:** $P_i(x|y) \neq P_j(x|y)$. Pedestrian images on different clients will exhibit the non-iid (Independent and identically distributed) issue, where the corresponding feature distributions are different clients.
- **Modality incomplete:** For an individual client, pedestrian images can fall into three scenarios: only visible, only infrared, or both visible and infrared.
- **Identity missing:** $P_i(y) \neq P_j(y)$. The subject is hard to appear in all cameras or entities.

Our Decentralized Privacy-Preserved Training (DPPT) Method

Preprocess: As mentioned before, existing VI-ReID methods adopt dual-stream architecture, which cannot be compatible with the CI protocol due to the challenges posed by **modality incomplete.** So we convert the two-stream structure into one-stream architecture and only sample images **according to identity rather than identity and modality.**



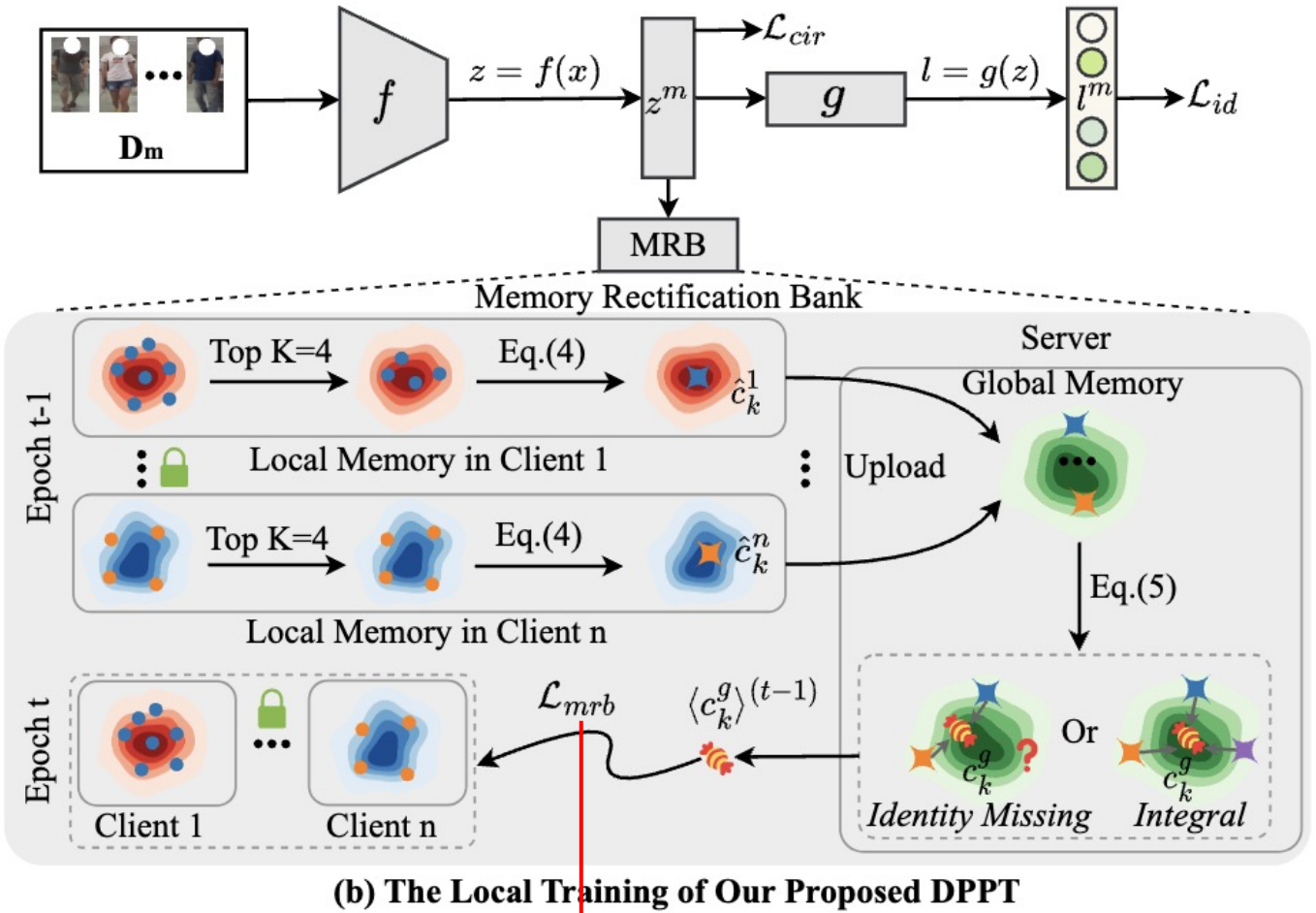
(a) Protocol of CI and EI

(b) The Local Training of Our Proposed DPPT

Figure 2. Protocols and proposed method. (a) are the protocols of camera independence (CI) and entity independence (EI) for privacy-preserved VI-ReID. Another our proposed protocol entity sharing (ES) is omitted in the figure, cause its training way is the same as that used in existing VI-ReID methods. (b) is the local training of our proposed decentralized privacy-preserved training (DPPT).

Tackle **domain shift** and **identity missing**

DPPT



(b) The Local Training of Our Proposed DPPT

$$\mathcal{L}_{mrb} = \frac{1}{N_m}\sum_{i=1}^{N_m}\left(1 - \frac{\langle z_i^m \rangle^{(t)} \cdot \langle c_{y_i}^g \rangle^{(t-1)}}{\|\langle z_i^m \rangle^{(t)}\| \times \|\langle c_{y_i}^g \rangle^{(t-1)}\|}\right)$$

Algorithm 1: DPPT

Input : The number of clients n, initial global model parameters θ^g, private local model parameters $\{\theta_1^l, \theta_2^l, ..., \theta_n^l\}$, local private datasets $\{D_1, D_2, ..., D_n\}$, and the number of training epochs E.

Output: The global model parameters θ^g for test.

for $e = 1$ *to* E **do**

  for $m = 1, 2, \ldots, n$ **do**

    $\theta_m^l, \mathcal{M}_m^l \leftarrow$ **ClientTraining**$(\theta^g, D_m, \mathcal{M}^g)$;

  // Upload and aggregate the local models

  $\theta^g \leftarrow \sum_{i=1}^{n} \frac{|D_i|}{\sum_{j=1}^{n}|D_j|}\theta_i^l$;

  // Upload and aggregate the local memory banks

  $\mathcal{M}^g \leftarrow \{c_i^g = \frac{\sum_{m=1}^{n} \mathbb{1}_{\{\hat{c}_i^m \neq 0\}} \cdot \hat{c}_i^m}{\sum_{m=1}^{n} \mathbb{1}_{\{\hat{c}_i^m \neq 0\}}} | i = 1, 2, ..., I\}$;

ClientTraining$(\theta^g, D_m, \mathcal{M}^g)$**:**

$\theta_m^l \leftarrow \theta^g$ // Download the global model to the local model

for $(x_i, y_i) \in D_m$ **do**

  $z_i = f_m(x_i), l_i = g_m(z_i)$;

  $\mathcal{L}_{cir} \leftarrow (z_i, y_i), \mathcal{L}_{id} \leftarrow (l_i, y_i)$;

  $\mathcal{M}_m^l \leftarrow z_i$ via Eq.(3,4);

  $\mathcal{L}_{mrb} \leftarrow (\mathcal{M}_m^l, \mathcal{M}^g)$ via Eq.(6);

  $\mathcal{L} = \mathcal{L}_{id} + \mathcal{L}_{cir} + \lambda\mathcal{L}_{mrb}$;

  $\theta_m^l \leftarrow \theta_m^l - \eta\nabla\mathcal{L}$;

return $\theta_m^l, \mathcal{M}_m^l$;

Comparison Experiment

Table 1. Evaluation of our DPPT under CI protocol. *Note that existing VI-ReID methods cannot directly be applied to the CI protocol as their frameworks are designed for data-shared learning.* Thus, we implemented four classic federated learning algorithms as baselines to verify the efficacy of our method: FedProx [19], Fednova [37], Moon [18], and FedAvg [23]. Evaluating metrics rank-1(%), rank-10(%), mAP(%), and mINP(%) are reported. AGW† and DNS† are the reproduced VI-ReID method that removes the modality information.

Methods	SYSU-MM01 [38]				RegDB [24]				LLCM [49]			
	r=1 ↑	r=10 ↑	mAP ↑	mINP ↑	r=1 ↑	r=10 ↑	mAP ↑	mINP ↑	r=1 ↑	r=10 ↑	mAP ↑	mINP ↑
FedProx [19]	25.90	68.83	27.48	17.18	25.62	44.19	25.96	17.80	23.72	53.95	30.59	27.69
+AGW† [44]	21.50	62.59	23.07	13.89	20.82	38.94	21.57	14.35	24.65	55.98	31.62	28.45
+DNS† [14]	36.11	78.29	35.22	22.02	46.75	69.27	43.06	28.99	26.35	57.00	32.88	29.54
+DPPT (Ours)	**38.16**	**81.54**	**38.15**	**25.17**	**51.33**	**71.43**	**48.93**	**36.15**	**27.46**	**59.33**	**34.60**	**31.44**
Fednova [37]	29.15	74.00	31.34	20.77	20.50	35.00	22.15	16.41	-	-	-	-
+AGW† [44]	22.00	64.15	23.52	13.54	13.83	25.47	15.92	12.21	-	-	-	-
+DNS† [14]	40.79	83.14	41.01	27.87	46.70	68.15	43.40	29.92	-	-	-	-
+DPPT (Ours)	**50.37**	**88.92**	**48.67**	**33.65**	**59.67**	**78.36**	**55.17**	**41.12**	-	-	-	-
Moon [18]	26.88	71.65	29.54	19.55	19.96	34.19	21.54	15.93	25.02	57.58	32.34	29.25
+AGW† [44]	20.79	62.57	22.44	12.80	11.19	20.56	13.44	10.46	23.60	55.98	31.62	28.45
+DNS† [14]	38.18	81.31	38.94	26.53	43.60	65.46	40.33	27.87	29.55	60.82	36.64	33.52
+DPPT (Ours)	**46.78**	**87.40**	**45.99**	**31.74**	**53.22**	**73.01**	**49.27**	**35.46**	**32.66**	**64.83**	**39.78**	**36.41**
FedAvg [23]	27.51	72.26	29.98	19.79	19.07	32.95	21.05	15.61	26.24	59.31	33.52	30.24
+AGW† [44]	21.65	63.13	23.25	13.45	14.17	23.84	15.89	12.10	24.31	54.51	30.66	27.19
+DNS† [14]	39.60	81.96	40.09	27.64	48.48	69.76	45.30	31.51	30.79	62.29	37.81	34.66
+DPPT (Ours)	**51.27**	**88.55**	**49.29**	**34.47**	**59.85**	**77.58**	**55.58**	**41.70**	**34.69**	**67.20**	**41.91**	**38.48**

We reproduced four classic FL algorithms and found that methods proposed after FedAvg even performed worse than FedAvg under the CI protocol. The reason is that these FL methods are **general methods** and **ignore the unique challenges** of VI-ReID.

Comparison Experiment

Table 2. Evaluation of our DPPT under EI and ES protocols. The upper part of the table is under ES, while the lower part is under EI. The underlined and bold indicate the best results for both protocols, respectively. B is the baseline using ERM with $\mathcal{L}_{id}$ and $\mathcal{L}_{cir}$ under ES, $B^{\dagger}$ denote the baseline using FedAvg supervised by $\mathcal{L}_{id}$ and $\mathcal{L}_{cir}$ under EI. We use R, L, and S to denote **R**egDB, **L**LCM, and **S**YSU-MM01 datasets, respectively. The left of $\rightarrow$ indicates seen entities and the right is the unseen entity.

Methods	Param.	FLOPs	R [24] + L [49] → S [38]				L [49] + S [38] → R [24]				R [24] + S [38] → L [49]			
			r=1 ↑	r=10 ↑	mAP ↑	mINP ↑	r=1 ↑	r=10 ↑	mAP ↑	mINP ↑	r=1 ↑	r=10 ↑	mAP ↑	mINP ↑
B	23.50	10.34	8.63	36.26	9.57	4.01	17.08	34.32	17.69	11.15	8.74	26.83	12.23	9.78
LBA [26]	23.55	10.36	8.09	34.63	9.55	4.29	12.01	28.65	11.69	6.01	8.38	26.80	12.20	9.90
AGW [44]	23.55	10.36	9.59	38.28	10.42	4.45	13.79	29.76	14.05	8.25	9.08	26.77	12.37	9.80
DEEN [49]	41.23	27.70	9.48	38.39	10.03	3.72	<u>19.42</u>	37.59	<u>19.44</u>	<u>12.33</u>	10.50	29.41	14.09	11.48
CAJ [43]	23.55	10.36	10.90	40.57	11.18	4.39	16.55	37.23	17.40	10.74	<u>11.35</u>	<u>31.17</u>	<u>15.03</u>	<u>12.10</u>
DNS [14]	25.45	10.36	<u>11.75</u>	<u>42.36</u>	<u>11.77</u>	<u>4.56</u>	18.87	<u>37.79</u>	18.36	10.56	10.14	28.47	13.55	10.94
$B^{\dagger}$	23.50	5.17	9.72	38.21	10.74	4.75	15.37	29.95	16.77	10.90	8.97	26.81	12.85	10.62
$B^{\dagger}$+CA	23.50	5.17	10.10	39.34	10.73	4.19	17.61	33.58	17.73	11.31	14.40	35.87	18.89	16.30
DPPT (Ours)	23.50	5.17	**11.27**	**41.20**	**11.86**	**5.34**	**21.54**	**40.37**	**20.72**	**12.78**	**14.63**	**37.03**	**19.15**	**16.44**

- While most methods slightly outperform our baseline B, the differences are minimal, and overall rank-1 accuracy remains low. This suggests that current VI-ReID methods still have limited capability in handling unseen environments.
- The baseline results under the EI protocol are similar to those under the ES protocol, indicating that **decentralized training does not significantly impact the model's ability to generalize to unseen environments.**

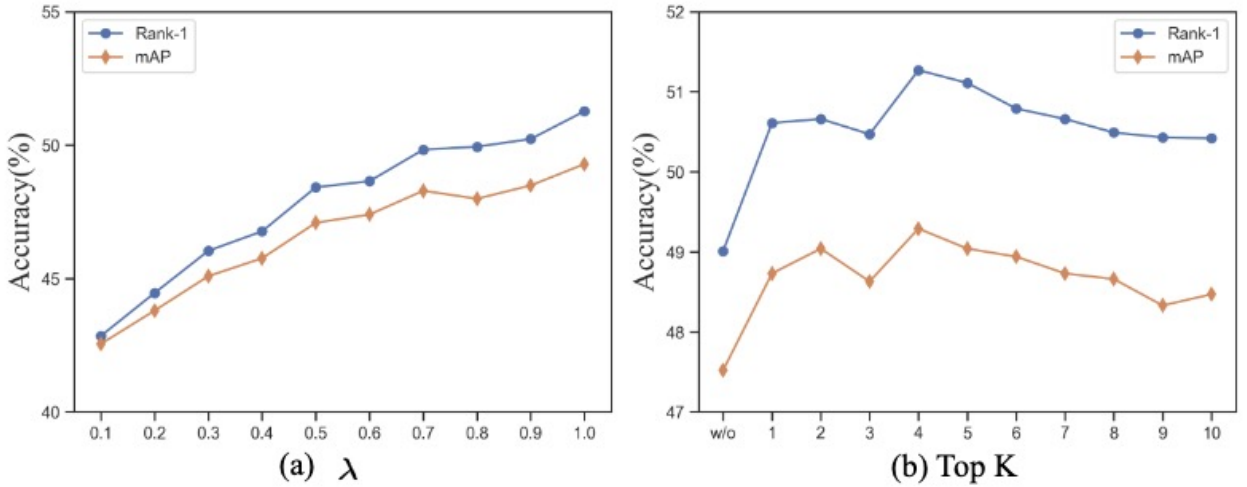
Experiments

Ablation study under the CI protocol:

Table 3. Ablation studies on SYSU-MM01 and LLCM datasets. The light purple is our final choice setting.

Settings	SYSU-MM01 [38]			LLCM [49]		
	r=1 ↑	mAP ↑	mINP ↑	r=1 ↑	mAP ↑	mINP ↑
B	27.51	29.98	19.79	26.24	33.52	30.24
B+CA	39.47	39.78	26.95	30.47	37.30	34.09
B+gray	35.25	36.16	24.03	29.33	36.29	33.06
B+CA+$\mathcal{L}_{mrb}$(cos)	**51.27**	**49.29**	**34.47**	**34.69**	**41.91**	**38.48**
B+CA+$\mathcal{L}_{mrb}$(edu)	43.79	44.00	30.77	33.45	40.62	37.35
B+CA+$\mathcal{L}_{mrb}$(mixed)	47.82	47.55	33.94	34.41	41.56	38.09

Analysis of MRB's Hyper-parameters:



Privacy evaluation:

Table B. Evaluation on various attacks.

Attack	r=1 ↑	r=10 ↑	mAP ↑	mINP ↑
image(100%)	46.36$^{\downarrow 4.91}$	87.32$^{\downarrow 1.23}$	46.53$^{\downarrow 2.76}$	33.21$^{\downarrow 1.26}$
gradient(33%)	42.82$^{\downarrow 8.45}$	84.99$^{\downarrow 3.56}$	42.57$^{\downarrow 6.72}$	29.10$^{\downarrow 5.37}$
gradient(66%)	32.20$^{\downarrow 19.07}$	75.00$^{\downarrow 13.55}$	32.74$^{\downarrow 16.55}$	20.95$^{\downarrow 13.52}$
gradient(83%)	24.00$^{\downarrow 27.27}$	65.35$^{\downarrow 23.20}$	25.32$^{\downarrow 23.97}$	15.12$^{\downarrow 19.35}$
gradient(100%)	4.90$^{\downarrow 46.37}$	22.76$^{\downarrow 65.79}$	6.10$^{\downarrow 43.19}$	3.16$^{\downarrow 31.31}$

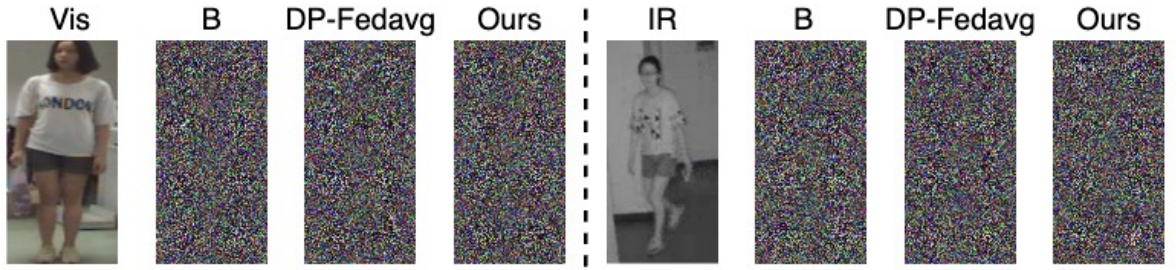
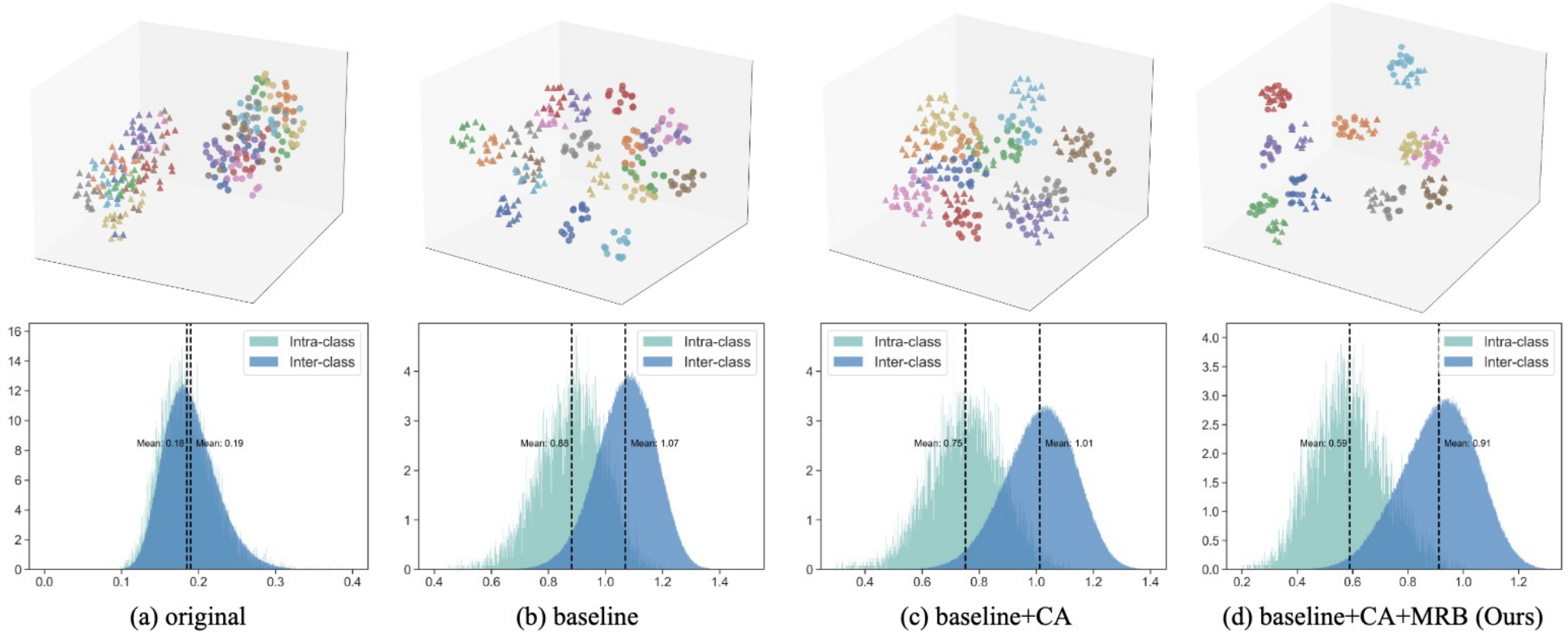


Figure A. Visualization of gradient inversion.

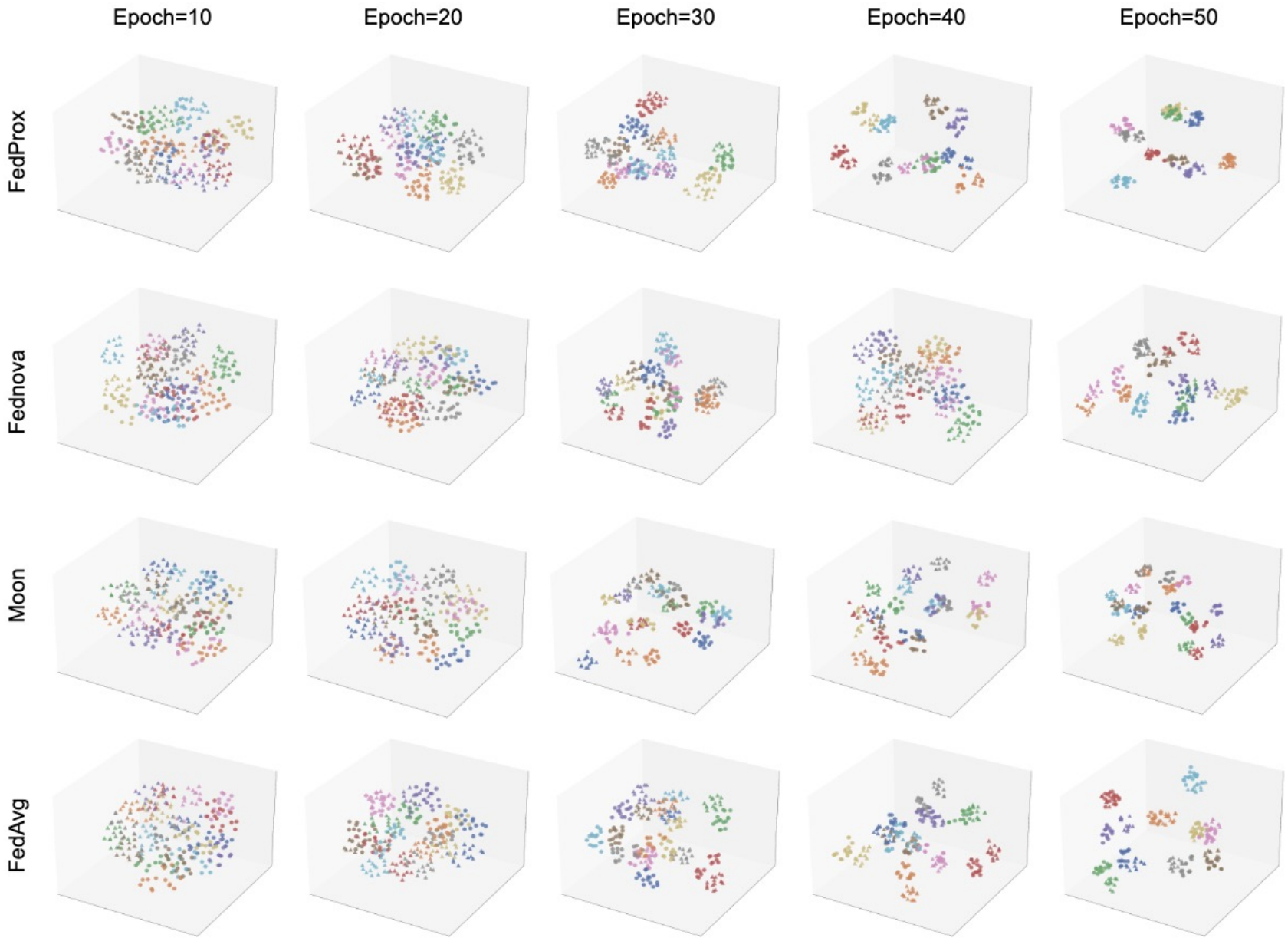
Visualization



(a) original (b) baseline (c) baseline+CA (d) baseline+CA+MRB (Ours)

Figure 3. t-SNE and intra-inter distances visualization of each component. The circles and triangles denote the visible and infrared modality, respectively. Different colors denote different identities.

Visualization



Conclusion

- We propose a benchmark named L2RW to simulate privacy-preserving real-world VI-ReID scenarios. Within the L2RW benchmark, we design three protocols, i.e., CI, EI, and ES, which simulates scenarios with different privacy constraints, bring VI-ReID evaluation closer to real-world conditions.
- We analyze the challenges encountered under the CI and EI protocols and propose a unified method named DPPT, which is the first work that handles privacy concerns for VI-ReID in a decentralized manner.
- In our L2RW benchmark, unlike existing methods that validate solely on a single dataset, we merge several existing datasets and conduct the evaluation in a cross-domain manner to simulate the real-world scenarios.
- Extensive experiments on three public VI-ReID datasets confirm the feasibility of decentralized training in L2RW, with our method achieving significant improvement on various federated learning baselines under CI. We also show that DPPT achieves performance under EI comparable to that of ES.

Thanks for watching!

Contact: yan.jiang@oulu.fi

Code